1. Requirements for Using GEM-TREND

GEM-TREND is implemented as a java applet which can be run in most common browsers and has been test with Internet Explorer 7.0, Internet Explorer 8.0, FireFox 2.0, Safari 3.2, Google Chrome 2.0 on different platforms. In some rare cases, the user may require to install JRE (Java Run-Time Environment) to run the GEM-TREND. JRE can be downloaded from SUN freely: http://www.java.com/ja/download/

2. Overview

GEM-TREND (Gene Expression data Mining Toward RElevant Network Discovery) is a web-based tool for retrieving gene expression data from Gene Expression Omnibus (GEO) by comparing gene-expression pattern of queries with those of GEO gene expression data and providing network visualization. The comparison methods are based on the nonparametric, rank-based pattern matching approach of Lamb et al. (Science 2006) with the additional calculation of statistical significance. GEM-TREND allows users find similar gene expression profiles without using keywords or IDs, and the results retrieved are not restricted by experimental conditions.

Retrieved gene expression data can then be viewed as a co-expression network with gene ontology (GO) annotation where genes and annotations are dynamically linked to external data repositories. The network analysis could provide insights into unknown functional relationships of the genes.

Figure 1. workflow of GEM-TREND
3. Gene expression data search

GEM-TREND provides both gene-expression pattern-based and text-based searches to retrieve gene expression data from GEO.

3.1 Gene-expression pattern-based search

3.1.1 Query preparing

For gene-expression pattern-based searches, the input data can be gene-expression signatures represented by a set of genes with up- or down-regulated labels or by gene expression ratio data. The format of the input data are shown as follows:

Format of gene-expression signature

Query of gene-expression signature is composed of two columns:
1) Gene identifier. This is the unique identifier for the gene. Recently, UniGene ID, AffyMetrix probe ID and Agilent Probe ID are accepted.
2) Gene label: assign 1 for up-regulation gene, -1 for down-regulation gene.

Note: GEM-TREND accepts maxium of 500 genes as signature.
Example is available at http://cgs.pharm.kyoto-u.ac.jp/services/network/sample/sample_signature.csv

Format of gene expression ratio data

Query of gene expression ratio data is composed of following two columns:
1) Gene identifier. This is the unique identifier for the gene. Recently, UniGene ID, AffyMetrix probe ID and Agilent Probe ID are accepted.
2) Log ratio of treatment to control:

Note: The number of genes is not limited, but only the top 500 differentially expressed genes will be automatically selected for search.
Example is available at http://cgs.pharm.kyoto-u.ac.jp/services/network/sample/sample_ratio_data.csv
3.1.2 Searching data using Gene-expression pattern (1) – Search –

1) Click “Browse” to upload the prepared query file (either gene-expression signature or gene expression ratio data). In this example, the query `sample_signature.csv` was used.

2) Select the species for search. The default species is humans.

3) Check box for calculating p-value.

4) Click “Submit” to submit query.

5) Click “Submit” to start searching.

Previous results of gene-expression pattern-based search and network visualizations can be retrieved by using the result ID and the network ID. IDs are available in two weeks.
3.1.2 Searching data using Gene-expression pattern (2) –Result–

GEM-TREND
Gene Expression data Mining Toward RElevant Network Discovery

- Check the box to select the record of interest
- Click the GSE ID or GPL ID to access GEO
- Results can be sorted by the header
- Both search results and selected series can be downloaded in CSV format
- The full series title can be displayed as a tool-tip when mouse is over the title
- Use JOB ID to retrieve current result later
- Click the “network” icon to view the gene co-expression network
- Department of Systems Biosciences for Drug Discovery, Kyoto University
3.2 Text-based search

GEM-TREND accepts keywords (e.g. p53), GEO series ID, GEO platform ID as query. Text-based search can be used to directly retrieved data from GEO, or used as further search to retrieve data from the results of gene-expression pattern-based search.

3.2.1 Retrieve data from the results of gene-expression pattern-based search

GEM-TREND accepts keywords (e.g. p53), GEO series ID, GEO platform ID as query. Text-based search can be used to directly retrieved data from GEO, or used as further search to retrieve data from the results of gene-expression pattern-based search.

If both keywords and ID are inputted, the “AND search” will be performed.
3.2.2 Retrieve data by directly searching GEO

For the data retrieved by directly searching GEO, the similarity score and P-value are not provided.
4. Network visualization

Retrieved gene expression data can be viewed as a co-expression network with gene ontology (GO) annotation where genes and annotations are dynamically linked to external data repositories. The network visualization page is composed of three major parts: the network graphical display area, the cluster information area, and the gene search window as showed in following figures. The network graphical display area dynamically shows the full or sub-network according to the user’s operation.

Mouse Manipulation at the network graphical display area

- Left single mouse click will select the node. The selected node centered sub-network will be displayed.
- Right single mouse clicking will invoke the pop-out menu. It allows user to expand the node, and hid the links.
- Move the mouse over the node will show the annotation information of the node as shown in the above figure.
- Mouse dragging changes network style.
- On the top of graph panel, there are two modes can be used to change network size (zoom mode) and to rotate network (rotate mode).
- The network will be refleshed after new operation. The network needs to be expanded by right click to invoke the pop-out menu.
4.1 Network with gene cluster information

The number showed in top-right of genes describes the number of hidden linkages. These linkages can be expanded or hidden by a right click on the gene of interest to choose from the pop-up menu.

Genes link to the UniGene database by clicking the UniGene icon or genes.

Click Cluster Name to view the sub-network which includes co-expression genes.

Genes (nodes) in red background are genes from query.

Genes in yellow background are the genes selected by user.

Network ID can be used to retrieve the network later. Network ID is available in two weeks.
4.2 Network with GO annotation

- Genes in the orange background are the genes associated with common GO term.
- Gene name is displayed as a tool-tip when mouse is over the node.
- The number following the term describes the number of genes associated with the term.
- Click the GO term to show the associated genes.
- Click the "GO" to jump to GO database.
4.3 Gene search in the network

Input gene ID or gene name to search a gene of interest in the network. The cluster which contains the hit gene will be shown under Gene Cluster tab, and the hit gene will be highlighted.